



# Le mouvement " Open Data " dans la grande transformation des intelligences collectives et face à la question des écritures, du web sémantique et des ontologies

Jean-Max Noyer, Maryse Carmes

## ► To cite this version:

Jean-Max Noyer, Maryse Carmes. Le mouvement " Open Data " dans la grande transformation des intelligences collectives et face à la question des écritures, du web sémantique et des ontologies. 2012. sic\_00759618

**HAL Id: sic\_00759618**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00759618](https://archivesic.ccsd.cnrs.fr/sic_00759618)**

Preprint submitted on 1 Dec 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Le mouvement « Open Data » dans la grande transformation des intelligences collectives et face à la question des écritures, du web sémantique et des ontologies

Jean-Max Noyer, Maryse Carmes, Grico - I3M, Université de Nice Sophia Antipolis - Prefics, Université Européenne de Bretagne - noyer@grico.fr, carmes@grico.fr

## I. INTRODUCTION

**Résumé :** Le capitalisme cognitif a deux caractéristiques principales, l'intelligence collective et l'utilisation intensive des technologies de l'information, fondées sur la numérisation du contenu, des procédures et des écritures. Dans cet article, nous essayons d'examiner le mouvement Open Data face à la transformation des intelligences collectives et analysons ce processus dans le cadre de l'action publique, de la science, de l'intelligence logicielle, de la stratégie. Nous montrons comment ce mouvement s'articule avec la question du web sémantique, des ontologies, avec la montée de l'algorithmique. Dans ce cadre, l'émergence du « data mining » se présente comme « récit impérial », comme le récit des sociétés performatives. S'ouvre également la possibilité de nouveaux modes de gouvernance, l'émergence de nouvelles façons de penser le politique et l'espace public.

**Abstract :** Cognitive capitalism has two main characteristics, collective intelligence and the intensive use of information technology, based on the digitization of content, procedures and writings. In this paper, we try to examine the Open Data movement faced with the transformation of collective intelligence. We analyze this process in the context of public policy, science, intelligence software. We show how this movement articulates with the issue of semantic web ontologies and with the rise of algorithmic. We emphasize, indirectly but strong, on the emergence of "data mining" as "imperial narrative," as the story of performative societies in the context of anthropological stratum Internet. We outline the possibility of new modes of governance and the emergence of new ways of thinking about politics and public space. The future of democratic societies is partly at stake

Mots clés : Cartographies, Gouvernance, Data Mining, Métadonnées, Indexation, Intelligences collectives, Open data, Open Archive, Ontologies, Web sémantique

D'un point de vue très général, l'Open Data (Données ouvertes) promeut l'idée d'un libre accès (et développe des dispositifs pour cela) à un certain nombre de données publiques afin d'en permettre une utilisation et une exploitation sans restrictions de droits d'auteur, de brevets ou d'autres mécanismes de contrôle. Il s'agit là d'une définition très simple. Le développement du web a favorisé, dès sa création, la mise à disposition "libre" de documents hétérogènes. C'est même l'idée fondatrice de ceux qui ont créé le réseau Internet sur la base d'un modèle spécifique de l'activité scientifique comme mode de production, circulation "libre" des savoirs, optimal en terme d'innovation et de créativité. On peut contester le côté irénique et simplificateur de ce modèle, mais c'est sur cette idée générale que le projet s'est développé.

La transformation numérique a pénétré tous les secteurs de la société, des sciences, de la santé, de l'éducation, du commerce, des entreprises, des institutions etc... et elle est amplifiée par la croissance extrêmement rapide de l'internet des objets. Ce que l'on appelle le « web des données » peut être vu comme un ensemble de procès qui convergent vers un objectif commun : la dissémination dans l'espace public et privé des données et la « réutilisation intelligente des données indépendamment de leur contexte numérique d'origine ». Cela pose aussi un certain nombre de questions et de difficultés, à la fois politiques, économiques et socio-cognitives.

De plus, alors que le Web se trouve travaillé, de manière intensive, par les phénomènes de différenciation et de fragmentation des ressources (et des conditions de leur production, de leur accès et circulation, de leur ré-exploitation), la tension se fait toujours plus vive entre, pour durcir la présentation, les tenants d'une philosophie ouverte des ressources (le web comme espace lisse) et les tenants d'une approche propriétaire (le web comme espace strié). Cette manière de présenter les choses est toutefois très imparfaite et comme le dit Moulher Boutang, offre trop de « facilités » conduisant à être chahuté « entre Charybe du marché pour tous de la propriété intellectuelle et Scylla des bantoustans du communisme primitif selon l'évangile de la nouvelle accumulation » [1]. Un grand nombre de données et ressources sont déjà librement accessibles sur le WEB bien que les conditions de leur réutilisation et exploitation soient variables, tant en terme économique-juridiques que socio-cognitifs.

Ce mouvement « Open data » se déploie encore alors que la marchandisation des données (pour ce qui est des données

numériques, déjà entamée dans les années 60, avec les grandes bases de données, numériques, bibliographiques, « full text », dans des domaines majeurs) ne cesse de croître rapidement et que la bataille techno-politique et juridique autour des applications, de la capture de données de plus en plus hétérogènes, s'intensifie.

Dans les pays les plus avancés de la transformation numérique, nous nous trouvons donc, entre mise en crise, affaiblissement des instances régulatrices et centralisées de l'Etat, transformation de l'espace public et montée en puissance de nouveaux modes de gouvernance plus décentralisée grâce auxquels la diffusion des conditions de la métastabilité de la société se fait par le milieu, se fait selon les modalités de type rhizome[2], et le déploiement d'un « Etat marché » où le traitement des données (en tous cas, de certaines) s'effectue sous les conditions de la privatisation des processus de création de la Valeur. C'est la raison pour laquelle nous sommes là dans un champ de bataille majeur.

De ce point de vue, il apparaît judicieux de situer notre réflexion dans la ligne du travail d'E. Ostrom, cette dernière montrant en effet que les « droits de propriétés font partie des arrangements institutionnels et que les formes de propriétés communes règlent la production et les usages des biens appartenant à des milieux complexes et fragiles de façon bien plus satisfaisante que la grossièreté du modèle unique de droit de propriété de l'individualisme passif de Locke sur lequel s'est édifié l'Homo Economicus ». [3]

Notre regard, posé ici sur l'Open data, vise à indiquer un certain nombre d'interrogations et de problèmes afin de mieux saisir la pragmatique interne des débats au sein des modèles de l'Open data vus par certains comme de puissants modes politiques alternatifs. Cette communication propose de réinscrire ces initiatives dans le mouvement des agencements au sein desquels ont été fondés des politiques de données et des modèles socio-techniques ouverts, distribués. Puis, en appui de divers milieux (collectivités, science, stratégie militaire), nous analysons un ensemble d'enjeux liés au couplage nécessaire d'une réflexion sur l'Open Data avec le web sémantique et les multiples dimensions du Data Mining. Ce faisant, nous établissons un ensemble de propositions pour ouvrir le champ politique et socio-cognitif de l'Open Data.

## II. LA FABRICATION DU MOUVEMENT OPEN DATA

### A. Brève histoire

La montée au premier plan des préoccupations de ce que l'on appelle les « Open Data » est une des conséquences de la croissance exponentielle des connaissances, informations, données, des savoirs produits, circulant, s'échangeant sur les réseaux numériques. Elle est un aspect de la lutte entre les nouvelles forces productives attachées à la connaissance et à l'importance croissante des moyens de pilotage sémiotique et, pour parler comme Marx, elle manifeste l'émergence de

nouveaux rapports de production.

Il y a cependant une très longue histoire « des données » comme instanciation, incarnation de la relation fondamentale qui lie savoir et pouvoir, où se joue la question du politique et de ses évolutions. Loin de nous le dessein d'en retracer ici l'histoire et les devenir. Mais, il n'est pas inutile de pointer quelques repères proches de nous, avec par exemple, au XIX<sup>ème</sup> siècle, la co-émergence de l'Etat-Nation et de la Raison Statistique. A. Desrosières a sur ce point produit un travail majeur.

Comme nous l'avons déjà noté, la question des données numériques elle-même prend une importance croissante dès la fin des années 50 et le début des années 60, avec le développement des grandes bases de données, scientifiques, commerciales, financières entre autres. Ont ainsi été rapidement conçues des bases de données stratégiques dans le secteur des publications de la Recherche, et ce à des fins de politique scientifique, de gouvernance scientifique.

C'est en effet au tout début des années 60 que les Etats-Unis se dotent des outils minima permettant d'évaluer les travaux des scientifiques, d'en suivre les dynamiques internes, d'en dégager les fronts de recherche... faisant apparaître les principaux enjeux liés à l'exploitation des grands fonds documentaires. Il en ira de même très rapidement, pour les bases concernant les entreprises, les brevets. La question des moyens de pilotage sémiotique à partir de la sphère numérique s'impose donc avec force et l'élaboration de méthodologies plus ou moins sophistiquées d'extraction de savoirs, va prendre là son essor sous les formes de la scientométrie puis de l'infométrie. Dès ce moment là, il est évident que l'on passe « du paradigme de l'accès au paradigme de la production de savoirs ». De plus, la montée en puissance de divers mouvements « Open » au début des années 90, dans le domaine de la création logicielle et de la dissémination des programmes, dans le secteur de la publication scientifique, constitue le milieu concret, expérimental, où s'exprime la remise *sur le métier* des questions politiques, organisationnelles accompagnant la variation plus ou moins agonistique des modes de production, circulation, exploitation, consommation des savoirs, des données.

Propulsés par les mouvements du logiciel libre, de l'Open Source et de l'Open Archive Initiative, les nouveaux modèles, contestant les dérives de l'extension apparemment sans limite du droit de propriété, ont tenté et tentent avec de plus en plus de force, de desserrer le lien dominateur entre les formes capitalistiques du droit de propriété intellectuelle, les dynamiques de l'intelligence collective, l'innovation et la création.

Les travaux de Lawrence Lessig, Richard Stallman, Steven Harnad, Paul Ginsparg sont ici séminaux [4].

En proposant des formes librement consenties de propriété collective, en s'appuyant sur la productivité des pratiques distribuées et sur l'adaptabilité de modes de gouvernance bottom-up, ces mouvements participent au changement positif des processus créatifs. Au plan anthropologique et politique, ils contribuent à la transformation de la culture délibérative démocratique, à l'apprentissage d'une culture de la controverse et à l'invention de nouveaux modes de gouvernance, polycentriques et variés.

Les conditions de production, circulation, exploitation des

connaissances, informations, données sont donc affectées et en débat. Les modèles économiques et politiques, mais aussi les régimes juridiques, subissent les assauts de nouveaux acteurs, de nouvelles demandes, pratiques, de nouveaux usages (GNU pour le logiciel libre, les licences Creative Commons et OpenDataCommons, en sont des actants importants).

Enfin, des technologies émergentes, rendant possibles de nouvelles manières « distribuées » de fabriquer des données, de les collecter et de les redistribuer dans les circuits des flux informationnels, sont en développement rapide, et ce, associées à la miniaturisation des interfaces, à la multiplication des applications logicielles de plus en plus sophistiquées.

Ainsi, dans l'ensemble de ce mouvement, les modèles de gouvernance centralisée et « top down » se trouvent devoir entrer en concurrence, co-exister, avec des modèles a-centrés, distribués et « bottom up ». Mais, en étant dans un rapport de co-détermination relative avec ces modèles, ce sont aussi d'autres batailles politiques qui s'ouvrent.

La compréhension du mouvement « Open Data » a tout à gagner à s'inscrire dans ces histoires courtes et longues, à ne pas rester suspendu dans le vide, un vide dans lequel il pourrait apparaître comme nouveauté radicale, sans milieu associé, sans « mémoires » ? Il ne s'écrit pas sur une table vierge. Il est pris d'emblée dans toutes les dimensions socio-politiques des rapports savoir-pouvoirs, il est pris dans les asymétries socio-cognitives individuelles et /ou collectives, dans les asymétries d'appropriation des technologies d'écriture, des technologies intellectuelles au sens large.

Et si, à l'occasion de l'irruption des mémoires numériques et des nouvelles technologies intellectuelles, se font jour des utopies nouvelles, concrètes ou pas, des désirs de transformation des économies politiques des savoirs, des projets et désirs de transformation dans l'art de créer de la valeur, des « valeurs », tout cela fait fond sur les hétérogénéités anthropologiques où s'éprouvent forces cognitives, forces créatives et désirs politiques.

#### *B. Actualisations du mouvement : gouvernements, science, stratégie militaire*

Dans un nombre important de pays, le mouvement « Open Data » prend de l'ampleur. Du gouvernement des Etats-Unis ([www.data.gov/](http://www.data.gov/)) au gouvernement de la France en passant par la Grande Bretagne l'ouverture des données publiques ne cesse de s'étendre. , Le 12 décembre 2011, « La Commission a adopté une stratégie européenne en matière d'ouverture des données qui doit soutenir l'économie de l'Union à hauteur de 40 milliards EUR par an ».

Ce mouvement se déploie de manière parfois paradoxale sur fond de la critique des Etats comme instance de régulation centralisée (Etat providence) et ce, d'une manière plus ou moins radicale selon que l'on examine la situation aux Etats-Unis ou en Europe, pour aller vite. Il est en effet intéressant de noter qu'aux Etats-Unis, ce mouvement se développe alors que l'Etat Marché tente de s'imposer de manière définitive. Tout ceci n'est pas sans importance dès lors que l'on veut bien considérer l'Open Data comme problème économique politique majeur, comme problème stratégique majeur. De

même, pour se limiter au paysage français, les régions, les départements, les villes, mettent peu à peu à disposition des publics, des citoyens, des ressources publiques variées, sous des conditions d'exploitation elles-mêmes variées [5]. Données de la science, données juridiques, données socio-économiques, mais encore, les systèmes d'adressage (comme des identificateurs géographiques qui expriment la localisation de lieux et événements) ou bien des données climatologiques, géologiques, des données liées à l'agriculture, à l'art... sont prises dans ce processus.

Les initiatives gouvernementales [6], nombreuses dans les pays démocratiques (et émergentes ailleurs), ne font que renforcer les expérimentations et réflexions sur les intelligences collectives, leurs modes de fonctionnement, leurs capacités réflexives, ainsi que leur capacité à déployer des modèles plus a-centrés (voire bottom-up) de production et circulation des savoirs... Ce sont là des questions complexes et décisives qui engagent à terme l'avenir des sociétés ouvertes et fluides. L'Open Data exprime la mise à disposition par les services publics et les gouvernements de certaines données publiques auprès des citoyens « pour qu'ils puissent par exemple les ré-utiliser dans des services et applications ». Les données liées à la démographie, à l'économie, au tourisme, au transport, à la culture, à l'environnement...etc peuvent ainsi être publiées et accessibles (selon des formats très divers). Tous les domaines de description et de gestion des territoires et des populations sont potentiellement concernés.

Et l'Etat et ses ministères, les institutions et entreprises publiques, les régions, les villes... sont convoqués à répondre de quelque manière que ce soit à ce mouvement et aux injonctions locales, gouvernementales et de la communauté européenne, tout en continuant à préserver le maintien de rapports équilibrés et protecteurs entre données personnelles, vie privée et « bien communs ». L'ouverture des mémoires numérisées, des bases de données (Open Data) du domaine public, le développement de nouvelles mémoires dans un nombre toujours croissant de secteurs de ce domaine public, doivent cohabiter avec la production de nouveaux types de mémoires dans le cadre de la société civile (ouvertes ou privées) et cela ne rend pas forcément aisée la compréhension des problèmes posés par ce mouvement qui se présente comme une célébration de la transparence comme vertu démocratique.

Cette célébration mérite, selon nous, une déconstruction urgente.

Plus encore, se développent ici et là, liées entre autres, aux problèmes du développement durable, de l'écologie, des nouvelles empiries numériques des sciences (Gérard Berry, Jim Gray) ... mais aussi de la sécurité, des « real-time environmental data bases » qui se fondent sur la collecte distribuée et la dissémination de capteurs dont les vecteurs sont des parties du public, des groupes de citoyens (exemple : les vélos écologiques dotés de capteurs de la Ville de Copenhague). Ce que l'on appelle le « Crowdsourcing » recouvre des processus très divers, diversité sur laquelle il conviendrait de s'interroger plus profondément. Rappelons

que le crowdsourcing est un processus qui vise à utiliser des sources et ressources nouvelles, non institutionnelles, ad hoc, participant à la production de connaissances voire dans certains cas à leur gestion. Ces sources étant impliquées tantôt dans des processus de capture et de collecte de données relativement simples voire passifs (état de capteurs consentants), tantôt dans des processus plus actifs de définitions des projets, de traitement et exploitation pouvant à leur tour être pris dans des processus d'innovation plus ou moins créatifs, plus ou moins liés à des conditions locales d'agencement collectif... Les formes d'implication des acteurs sont là encore variées, du bénévolat à des formes rémunérées, elle peuvent aussi prendre des formes collaboratives plus ou moins complexes...

Dans le cadre de la vulgarisation scientifique et des prétentions à intégrer de plus en plus de citoyens dans la prise de conscience du processus complexe de co-construction des sciences, la constitution des dits citoyens comme vecteur de la collecte (voire d'une production plus élaborée) tend à se développer : les cas de l'astronomie, des sciences environnementales sont bien connus de tous. Les effets de cela, sur la diffusion de la culture scientifique et le développement (du point de vue de la science) d'une réflexivité accrue, sont parfois difficiles à saisir et ce d'autant plus que la montée de formes de controverses socio-techniques associée à la crise des rationalismes, est en cours.

Selon une perspective proche, les relations entre les experts et le monde « profane » se trouvent aussi être en voie de transformation. C'est le cas par exemple des mouvements de malades qui veulent partager les connaissances avec les médecins, faire entendre leurs voix et leurs approches des maladies, les médecins pouvant à leur tour profiter, si l'on ose l'expression, de « l'expertise profane » et des données issues de programmes de collecte contrôlée... On pourrait multiplier les exemples et les expérimentations. (Smart Cities et Urbanisation, Ecologies, Mégapoles...)

Nous mentionnerons cependant encore ceci, afin de bien montrer l'étendue des questions portées par l'Open Data. Sont aussi concernées la stratégie et l'utilisation des Open Data et plus généralement des sources ouvertes et du Data Mining, comme éléments de la lutte contre-insurrectionnelle. Dans le cadre d'interventions extérieures, l'incitation à la mise en place de réseaux numériques, de dispositifs communicationnels numériques et de bases de données locales, tend à constituer un axe majeur à la fois stratégique et tactico-opérationnel au sein de la sphère sémiotique et ce, afin de mieux maîtriser les dimensions anthropologiques, culturelles, sociétales du conflit ou de l'intervention en cours.

C'est ainsi que le rapport « Byting Back. Regaining Information Superiority Against 21st-Century Insurgents » poursuit l'exploration des moyens de contrôle des dimensions politiques et anthropologiques des champs conflictuels et insurrectionnels à partir des dispositifs communicationnels. Ces derniers s'appuyant sur les mémoires culturelles et politiques de l'espace numérique, les « Open Data », et sur les gisements de savoirs collaboratifs que constituent les réseaux sociaux numériques de toutes sortes et en constant développement [7].

De même la multiplication des Open Data et leur exploitation entre en résonance avec la volonté sécuritaire croissante et le développement de l'urbanisme militaire.

*Suivre à la trace le citoyen-consommateur-soldat est en effet au cœur des sociétés de contrôle et de veille. D'une manière générale « négligées et considérées comme allant de soi – aussi longtemps en tous cas qu'elles continuent de fonctionner—les infrastructures qui assurent le fonctionnement de la vie urbaine se retrouvent de plus en plus au cœur de la violence politique et de la doctrine militaire contemporaine. La collecte de données impliquant forces militaires et policière aidée par la robotisation en cours et l'utilisation d'une algorithmique de plus en plus sophistiquée [8].*

### C. L'ouverture du champ politique

Le mouvement « Open Data » est l'occasion d'une renégociation politique des relations entre savoirs et pouvoirs, d'une transformation des intelligences collectives, d'une libération des flux de savoirs qui circulent. C'est aussi l'occasion d'élaborer de nouvelles critériologies afin d'évaluer son impact économique et politique et ce, dans le contexte des évolutions des écologies environnementales, du travail, du psychisme, de la culture elles-mêmes sous les conditions et contraintes des marchés capitalistiques.

Le mouvement « Open Data » serait ainsi plus crédible encore, s'il prenait en charge la question de l'ouverture des accès à des ensembles de données à partir d'un point de vue réaliste des usages, des pratiques socio-cognitives, des économies politiques et des contraintes attachées à la production de nouveaux savoirs et de connaissances.

Plus que jamais aujourd'hui, fractures numériques et fractures cognitives entrent en résonance et une agonistique des narrations [9], des forces de création est engagée y compris à partir du champ d'immanence doxique pour reprendre l'expression de Philippe Mengue. [10],

Les usages et les écritures de et à partir des données numériques sont donc extrêmement différenciés, les capacités d'extraction de savoirs nouveaux à partir de ces données, la remise en jeu de ces savoirs dans le processus de circulation et de production etc ... sont soumis à des cribles très discriminants. Et il apparaît de plus en plus nettement que l'Open Data ne pourra faire l'économie d'une réflexion sur la formation citoyenne à la traversée du monde sociétal par l'informatique, par la numérisation.

Ce mouvement devra aussi se poser la question de son articulation avec la dissémination des technologies intellectuelles, des logiciels même les plus élémentaires, visant la manipulation intelligente de ces données ainsi que la créativité individuelle et collective. Il se trouve donc devoir faire face à la question du nécessaire développement du web socio-sémantique, à la création d'interfaces démocratiques et à

leur apprentissage.

*D. Ichnologie et Interfaces : le tissage continu des liens et des données*

Le déploiement des écritures et des réseaux numériques entraîne une transformation des mémoires, des dispositifs de publication, des niches écologiques qui sont conditions de notre vie, de nos modes d'existence et ce, dans leurs dimensions publiques et privées. Ce déploiement est planétaire et il affecte, bien que de manière très différenciée et inégale en terme d'actualisation, nos anthropologies, nos pratiques sociales, nos pratiques sociocognitives, nos subjectivités, nos processus d'identification. De nouvelles économies politiques émergent associées à de nouvelles économies libidinales. A partir des nouvelles dimensions réticulaires, une nouvelle onto-éthologie des individus est en cours de déploiement et les collectifs de travail et de pensée sont profondément recomposés. N'en déplaisent à certains, nous ne sommes pas confrontés à une perte de liens, à un appauvrissement des associations, de ce qui nous attache et nous lie, mais à une transformation, à une différenciation conduisant à une saturation de tout cela où se joue pour une part essentielle la question de l'avenir de nos devenir et des processus d'altération, des sociétés ouvertes. Le monde apparaît de plus en plus comme un vaste ensemble de systèmes relationnels où de nouvelles cartographies sont nécessaires pour habiter son intérieur. Cela résonne avec une évolution plus ou moins accentuée du « Politique » comme expérimentation généralisée et comme domination problématique des moyens et de la performativité des procédures sur les Fins éthiques et politiques. Les grandes crises écologiques (sociale, mentale, environnementale) qui sont à la traversée de transformations anthropotechniques, économiques, cognitives, organisationnelles, militaires etc. pour partie majeure, sous les conditions du numérique et les devenir scientifiques et techniques associés, accentuant ces tendances et processus. Devenir biopolitiques et psychopolitiques, prenant des formes nouvelles, inédites.

On comprend alors toute l'importance des efforts qui tendent à développer une sémiopolitique des interfaces [11] capable d'indiquer les rapports de production nouveaux, les formes nouvelles d'asservissement (au sens machinique de ce terme) et d'assujettissement [12], de leur hybridation, qui se mettent en place par et au cœur des dispositifs informationnels communicationnels numériques, sans parler des formes existentielles, des temporalités et des trouées dans la texture étouffante des sociétés de veille. Texture qui se trame à partir d'une ichtnologie [13] toujours plus extensive et intrusive, le spectre de la traçabilité hantant les nouveaux modes de travail, les nouveaux modes d'existence et les nouveaux modes de gouvernamentalité.

Cette stratégie généralisée des interfaces et donc des normes, constitue un des piliers majeurs de la transformation de la gouvernamentalité aujourd'hui, plus précisément dans le cadre du devenir « Empire » tel qu'il est défini par A. Negri et M. Hardt[14]. On assiste en effet à un affaiblissement relatif des systèmes de contrôle centralisé, et à un renforcement des systèmes de contrôle, de veille, immanents au système de production des réseaux distribués. La gestion de ces vastes et complexes réseaux a-centrés se faisant sur un mode multifractal, avec un grand nombre de boucles récursives et de règles locales, ces dernières portées par des interfaces machiniques numériques de plus en plus sophistiquées.

Quels que soient les niveaux d'échelle, des intranets aux processus de globalisation avec les passages en cours de la forme de l'état nation, vers l'état marché (dans le cadre de la tentative d'établissement, à marche forcée, d'un marché mondial), des formes classiques de la souveraineté vers des nouvelles formes décentralisées, relativement complexes (ou la privatisation même des fonctions régaliennes va bon train y compris la privatisation des machines de guerre), ces passages donc, se manifestent encore à travers la montée en puissance de la question logicielle et sémiotique comme question politico-stratégique majeure. Et qui peut extraire et exploiter les graphes correspondants à partir des traces numériques (singulières et / ou collectives) occupe une position de supériorité au sein des économies politiques, libidinales, stratégiques. Qui peut exploiter les variations des rapports de vitesse et de lenteur entre les systèmes d'écritures, les variations des combinatoires entre « memes »... opérant au cœur de la question socio-politique des modes d'intelligibilités et des processus de subjectivation, acquiert une position dominante dans la production circulation, consommation des savoirs.

### III. L'EXPLOITATION DES DONNEES: WEB SEMANTIQUE ET DATA MINING

#### A. Le débat épistémologique

Dans le cadre posé précédemment, ce qu'on appelle, les Open Data, les Big Data, les Small Data... se trouvent donc au centre de tensions socio-cognitives et politiques majeures, mais aussi au sein de débats épistémologiques importants, en particulier ceux portant à la fois sur la puissance de l'empirie numérique et de l'algorithmique statistique, sur la remise en cause de la place de la Théorie, sur l'efficacité de ses dimensions spéculatives et de la modélisation conceptuelle. [15] Tel que l'indiquait Chris Anderson en 2008, "At the petabyte scale, information is not a matter of simple three- and four-dimensional taxonomy and order but of dimensionally agnostic statistics. It calls for an entirely different approach, one that requires us to lose the tether of data as something that can be visualized in its totality. It forces us to view data mathematically first and establish a context for it later ». (...) « There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might

show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot. » [16].

D'autres interrogations sont à l'œuvre dans le champ des Sciences Humaines et Sociales. Il convient de bien mesurer les évolutions des écritures numériques en particulier sous la forme de ce que l'on appelle, le web sémantique et socio-sémantique, ou bien encore la question des ontologies. [17] Ces écritures sont là encore en première ligne et la maîtrise et la conception des mémoires numériques participe pleinement de ce que Jean-François Lyotard nommait en 1979, " la légitimation par la puissance. (...). Cette dernière s'autolégitime comme semble le faire un système réglé sur l'optimisation de ses performances. Or c'est précisément ce contrôle sur le contexte que dit fournir l'informatisation généralisée. La performativité d'un énoncé, qu'il soit dénotatif ou prescriptif, s'accroît à proportion des informations dont on dispose concernant son référent. Ainsi l'accroissement de la puissance et son autolégitimation, passe à présent par la production, la mise en mémoire, l'accessibilité et l'opérationnalité des informations". [18] Mais pour suivre ici Yannick Maignien, [19] la révolution numérique est à la fois l'objet et l'instrument du développement contemporain des SHS.

#### *B. Collectivités territoriales et méta-données : un horizon (trop) restreint*

Face à l'exploitation des données, à la production de nouveaux savoirs, à de nouvelles formes d'expression (y compris et surtout) grand public et d'accès, à leur libre utilisation (socio-politique), les diverses formes du mouvement Open ont dû affronter la question du traitement des données, des corpus documentaires, des traces... Les problèmes posés sont de nature cognitives et à ce titre ils s'inscrivent dans le travail scientifique. Ils sont encore de nature politique et économique car la production de « Valeur » a des coûts.

On voit aujourd'hui par exemple, au cœur du mouvement de l'édition « Open », mais aussi dans le cadre de l'édition privée, monter la question de « l'Open Content Mining ». [20]

Nous avons montré ailleurs [21] les liens étroits qui unissent les mouvements de l'Open, de l'Open Data et le Data Mining, lui-même au centre du processus de marchandisation des données, des savoirs et connaissances, des sémiotiques, dans les régimes des économies de marché. Si l'on observe, par exemple les collectivités territoriales françaises, elles abordent la question de l'Open Data à partir de leur propre héritage institutionnel et à partir de leur posture « top down ». Dans ce cadre, elles n'expriment qu'une des manières de penser l'Open Data comme problème politique, économique, cognitif, culturel... Il s'agit pour elles, principalement, de traiter l'accès aux données, leur distribution en laissant provisoirement ( ? ) la question des métadonnées attachée aux

besoins de d'interopérabilité institutionnelle et organisationnelle. [22]

Les métadonnées sont pensées selon le paradigme de l'accès et de la performativité interne de l'agencement des collectivités entre elles, sans que soit pour l'instant « penser » la question socio-politique de l'innovation, la question du renouvellement de l'exercice démocratique. Des conditions d'écritures, d'interfaces, de politique d'ingénierie documentaires sont là à l'œuvre. Pour ces collectivités, ce qui prime pour l'instant et qui constitue un des principaux axes de travail, consiste donc en une *normalisation des catalogues de données des portails d'Open Data*. La proposition portée par elles, n'est jamais qu'un équivalent tardif (et nécessaire) de ce qu'a été le Dublin Core pour le Web. Il s'agit de mettre en place une fusion des différents catalogues selon le format DCAT (Data Catalog Vocabulary) du W3C. [23]

Mais, l'évolution est rapide et l'attention devrait se porter rapidement vers des questions du type : "que faire", "pourquoi faire" et pour quelles finalités, ces méta-données ? Interconnection(s) et traitement(s) intelligents des données et des corpus deviennent essentiels. Cela suppose que les conditions de la création logicielle, de la création des applications de leur dissémination soient elles-mêmes abordées de manière ouverte. La tension là encore est grande entre modèles libéraux (voire néo-libéraux) et les divers manières d'envisager les commons. Et ce, d'autant plus que les modes de création et de sélection des usages et les désirs d'intelligence sont non seulement variés mais parfois en compétition, en conflit. Concevoir des « commons » sous les conditions de la multiplicité ou bien sous les conditions de l'unité revient à fabriquer des modèles de gouvernance soient a-centrés soit centrés. Les travaux portés par E. Ostrom sont encore ici, référence forte [24].

La question des métadonnées et celle du Web sémantique doivent donc être conduites sans préjuger des formes d'organisation plus ou moins optimales (selon quelles critères et fins ?) et examiner les divers modèles de web sémantique ou socio sémantique et ce qu'ils autorisent. Rien ne serait pire en effet que d'essentialiser les écritures et les approches, de développer des ontologies fermées, au moment même où les besoins de gouvernance doivent viser la processualité et l'hétérogénéité des acteurs et des pratiques, des critères et des fins, doivent avoir pour but d'ouvrir aux rapports différentiels entre micro-mondes et mondes. Ce que signifie « réutilisation des données » c'est bien la relance du procès de travail, de la lutte pour de nouvelles formes démocratiques, des nouvelles formes de créativité. C'est une vaste tâche.

Pour aller, dans un premier temps à l'essentiel, les standards principaux publiés par le W3C s'appellent RDF, OWL et SPARQL. Et un certain nombre d'outils ont été développés à partir d'eux. Les approches fondées à partir du schéma RDF et de sa forme prédicative, sont une des manières (parmi d'autres) de concevoir certaines exploitations des données (Linked Open



Data” ou “LOD cloud”), de concevoir les traitements divers d’un même corpus de données. Les points de vue sur le monde sont multiples et non nécessairement convergents et cela devrait suffire à intriguer les concepteurs d’ontologies, de métadonnées, à éviter la tentation la mieux partagée du monde de créer écritures et standards à partir d’une position totalisante et réductrice.

C’est une des raisons pour laquelle il convient de suivre avec attention les travaux du « Government Linked Data Working Group » qui a pour mission « de fournir des standards et des informations qui pourront aider les gouvernements dans le monde à publier leurs données, sous forme de « Linked Data » efficaces et fonctionnelles grâce aux technologies du Web sémantique ». [25]

Il en va de même pour « SKOS ». Le programme SKOS (Simple Knowledge Organization System) définit un modèle de données pour partager et combiner les systèmes d’organisation des connaissances par l’intermédiaire du Web.

### *C. Data Mining : des métadonnées aux ontologies, la création de Valeur*

Dès le début de l’approche « Open », nous l’avons déjà indiqué, le passage du paradigme de l’accès au paradigme de la production-circulation des savoirs sous des conditions nouvelles, a mis en lumière l’importance majeure des métadonnées afin d’accroître la qualité et la puissance des inférences possibles pour extraire des savoirs, des modes de traitement, par exemple statistiques, permettant d’exploiter les données hétérogènes, structurées, semi-structurées ou non structurées... Nous verrons que ce que l’on subsume sous l’expression Data Mining, réutilisation et exploitation des données, pour la création de nouvelles visibilité, pour la création de valeur est ici essentiel. C’est la raison pour laquelle nous pensons que nous ne pouvons traiter séparément dans le cadre d’une réflexion générale sur l’Open Data, les multiples dimensions du Data Mining, des pratiques intellectuelles et en particulier des conditions socio-cognitives des usages, des citoyennetés, des consommateurs, des élites... Individus, organisations, communautés, font face à l’ouverture des données selon les dispositifs socio-cognitifs qu’ils portent, fabriquent, selon les finalités politiques, économiques, éthiques qu’ils se donnent.

Cela nous amène à saisir encore une fois, « les limites de la situation actuelle : un web axé sur les données et une coupure nette entre le champ des sciences sociales et celui des nouvelles technologies ». Pierre Levy a pour sa part envisagé, depuis longtemps, de refonder ce lien à travers un métalangage (Ieml) qui doit permettre de créer les conditions d’une navigation « intelligente » entre des domaines et des ontologies hétérogènes, en tous cas, de permettre d’habiter des zones de traduction et de contact où les exercices cognitifs, le travail de la pensée pourront s’effectuer. Il s’agit là de développer donc les conditions de réflexivité, nécessaire à l’utilisations des collectifs de pensées, comme instance de leurs propres opérations. [26]

Dans ce contexte et dans le contexte élargi de l’économie politique du marketing, de l’économie sécuritaire, de l’urbanisation et du développement durable, de l’extension des moyens de pilotage sémiotique... *le Data Mining est devenu un nouveau “grand récit”, récit souterrain et pourtant aveuglant qui tend à structurer et à façonner le monde.* Des Sciences Humaines en passant par le Marketing et l’Economie, les dimensions Sécuritaires et Stratégiques, les diverses formes de Data Mining et modes d’intelligibilité, ont fait se nouer en des alliances parfois troubles et dangereuses, techniques de profiling et prédiction, performance et préemption. [27]

Plus que jamais, la tentation, qui d’un certain point de vue pourrait être funeste, de produire un contrôle continu de la réalité politique, de la réalité stratégique et économique, de la psychopolitique, trouve là de quoi nourrir ses rationalités insomniaques. Le Data Mining étant la colonne vertébrale assurant le passage des sociétés de contrôle aux sociétés de veille ou de “sousveillance”.

Quelle est la position de ceux qui parlent et agissent au nom de l’Open Data?

### *D. Le modèle limité mais fondateur de L’Open*

Ce qui est porté par exemple, par le mouvement de l’Open Archive, nous oblige à penser les relations entre ce que l’on nomme (comme diverses facettes des transformations du procès de travail intellectuel, des modes de gouvernance et des économies politiques associées aux intelligences collectives) Open Data, Big Data, Linked Data, e-Science, Data Mining, Algorithmique...

Ces rapports sont traversés par un double effort. Comment améliorer, structurer ces réservoirs de données (traces diverses, comportementales, documents de toutes sortes, données scientifiques, financières...) d’un côté, afin d’en extraire de nouveaux savoirs et connaissances, et de l’autre comment exploiter les données semi-structurées ou non structurées pour faire de même ?

Comment donc gérer les énormes quantités de données hétérogènes, en tirer des connaissances et des savoirs, et comment rendre ces analyses et leurs résultats réutilisables par d’autres acteurs? On sait les difficultés à concevoir des approches, des langages combinatoires permettant de créer des « ontologies » telles qu’elles puissent entrer dans des rapports créatifs. La co-existence des diverses manières de modéliser les comportements socio-cognitifs et les contenus utilisés et produits associés est à n’en pas douter, un enjeu complexe qui met à mal les tentatives globalisantes et les efforts pour créer les conditions d’une combinatoire universelle (IEML[28]par exemple en est une des tentatives les plus marquantes). La modularité semble pour l’instant de mise et les approches singulières, les plus efficaces, avec comme horizon la multiplicité des modèles...

### *E. Un changement de paradigme ?*



Nous pensons que les tendances portées par la variété et la différenciation des Open Data et leur mise au centre des économies politiques des savoirs, des intelligences collectives, doit conduire à un changement de paradigme dans le champ des ontologies et / ou onto éthologies, dans les conditions de leur production. Pour des raisons liées au maintien des sociétés ouvertes, à la possibilité de pouvoir mener des controverses de tous ordres, nous pensons qu'il est souhaitable de ne pas seulement s'enfermer dans le cercle gris des métadonnées primaires et d'ontologies trop globalisantes, conçues massivement à partir de logiques prédictives, essentialistes dans la postérité des limites et des rêves éveillé de l'intelligence artificielle symbolique. [29]Déployées de manière top-down sous des conditions et des intérêts d'un petit nombre de groupes et collectifs elles ne peuvent prétendre à dire comment doit être le monde, même si elles sont un passage en partie obligé. Le développement d'ontologies locales liées aux développements des applications, au développement logiciel et algorithmique à partir d'une posture bottom-up et associées à des données d'observations variées ne cessent d'apparaître comme essentiel et vital pour faire face à la complication du monde et aux hétérogénéités qui le parcourent et le fabriquent. Les prochaines générations d'infrastructures cognitives, c'est-à-dire les technologies intellectuelles (écritures, interfaces, cartographies...) doivent (devraient) permettre à une partie des utilisateurs de devenir acteurs plus ou moins puissants au cœur des agencements cognitifs eux-mêmes.

Le Web sémantique et socio-sémantique, consiste à concevoir la question des métadonnées selon les inférences et les exploitations de données que l'on souhaite développer, selon les programmes de recherches, les usages que l'on a en vu. Ce que l'on appelle « la réutilisation » doit donc être compris de manière très extensive à savoir : utilisation dans d'autres contextes, exploitation, invention de nouvelles données, de savoirs nouveau, encore une fois, pour une création de Valeur.

Des métadonnées aux fabrications (jusqu'aux générations semi-automatiques ou automatiques) d'ontologies et à leur interopérabilité partielle, ce qui se trame et est en jeu, c'est non seulement la création de Valeur au sens évoqué précédemment, mais aussi différentes façons de concevoir le travail intellectuel, de concevoir et pratiquer les problèmes liés aux sociétés ouvertes et démocratiques.

Dans quelle mesure cette normalisation peut être dépassée ou en tous cas, laissée ouverte pour l'introduction de métadonnées productives de nouveaux savoirs, il y a là à n'en pas douter un champ de bataille du point de vue de la question démocratique et de la liberté de penser. De ce point de vue, les mouvements de l'Open Source et de l'Open Archive Initiative permettent de se faire une idée des enjeux politiques, économiques, scientifiques. Approche par les « Commons », approche par les APIS propriétaires, hybridation des deux, sont donc en débat.

#### *F. Open Data et modes sociaux de classification*

Nous devons avoir à l'esprit que le problème des métadonnées, dans le cadre de l'Open Data, ne peut éluder la question de ses relations avec le développement des modes sociaux de classification : indexation sociale et doxique. De quelle nature sont ou seront ces relations ? Quelles possibilités seront données aux utilisateurs d'introduire leurs métadonnées, selon quelles procédures et liberté ?

En effet, l'évolution du web est caractérisée par la possibilité offerte à tous de pouvoir éditer, annoter, classer, lier... Le « bookmarking » social ne cesse de croître ainsi que les dispositifs ouverts de « tagging », plus connus sous le nom de folksonomies. Ces dernières représentent une grande source potentielle de richesse sémantique. Il a été vite perçu que cette approche émergente et ascendante (ou « bottom up »), des ressources collaboratives pouvait contribuer à surmonter certaines difficultés rencontrées dans la gestion des savoirs. Plus délicate qu'il n'y paraît au premier abord, l'exploitation des folksonomies, couplée aux technologies du Web Sémantique, est pensée à présent comme un élément essentiel pour le passage du « Web 2.0 » au « Web 3.0 ». Ces problèmes et ces recherches relativement complexes ont pour but la création de nouveaux modes d'indexation et de description. La production de métadonnées, y compris les folksonomies, est un enjeu majeur pour ceux qui souhaitent développer une conception du Web de type socio-sémantique et influencer sur le rapport entre un « Web de documents » et un « Web de données ». Plusieurs axes de recherche se déploient en parallèle, dont nous donnons ci-après des exemples.

Dans le cas d'OAI-ORE (Open Archives Initiative Object Reuse and Exchange), le but est de fournir « les standards pour la description, l'échange d'agrégats de ressources Web. Ces agrégats autrefois appelés "objets numériques complexes" (compound digital objects), peuvent associer des ressources distribuées composées de textes, d'images, de vidéos, de données. La finalité de ces standards est de rendre visible le contenu riche de ces agrégats, et utilisable par des applications qui prennent en charge le dépôt, l'échange, la visualisation, la réutilisation et la conservation » [30]. Cette création des métadonnées par les usagers intéresse fort les communautés dites « Web 2.0 » et les tenants du web sémantique et socio-sémantique. L'intérêt du « tagging », qui s'appuie sur des dispositifs informatiques donnant la possibilité à l'utilisateur d'indexer des documents, est qu'il facilite l'organisation et l'exploitation de ses propres ressources. Il s'agit d'un mode plus ou moins informel d'indexation et de classification distribuée.

Rappelons que les annuaires internet et les folksonomies se développent à partir de communautés d'indexeurs coopérant à travers le web. Les annuaires de ressources s'inscrivent dans la logique des schémas de classification de la bibliothéconomie, sans toutefois revendiquer l'organisation disciplinaire et la légitimité du champ théorique auxquelles ceux-ci se conforment. « Les annuaires de ressources internet qui sont construits par des communautés de bénévoles ne cessent de se

développer. A la différence d'un annuaire comme celui de Yahoo, par exemple, qui met en jeu des salariés placés sous l'autorité de managers définissant une politique de référencement explicite, le processus social d'indexation mis en oeuvre dans ces communautés relève d'un type de contrôle social situé entre celui de Wikipédia et celui, plus hiérarchique, des communautés "Open Source". Selon leur niveau de réputation, les membres de la communauté virtuelle peuvent contrôler des niveaux plus ou moins élevés de la classification, rajouter de nouvelles branches dans les domaines au sein desquels ils font "autorité", supprimer les descriptions effectuées dans la "notice" par des participants moins renommés, etc. » [31] Les avantages et les inconvénients des folksonomies ont été très discutés. L'activité de « *tagging* », en tant que moyen doxique d'indexation (en partie intuitif), favorise un renforcement des pratiques coopératives entre acteurs. Toutefois, l'absence d'une organisation *a minima* des « *tags* » et leur hétérogénéité sont des points de faiblesse qui, entre autres, ne permettent pas une re-exploitation facile et inter-communautaire. C'est la raison pour laquelle il convient, non seulement de définir des standards, mais encore de proposer des méthodes rendant possible l'exploitation et le partage des ensembles de « *tags* ».

Plusieurs approches sont utilisées ou sont en cours de développement. Elles sont fondées, soit sur des approches sémantiques et sur l'établissement de règles formelles (type Sujet-Verbe-Prédicat : le schéma Rdf), soit sur des approches statistiques-linguistiques inspirées par les travaux de la linguistique mathématique ou par les techniques de clusterisation (issus par exemple de l'infométrie)[32]

Le programme SCOT (*Social Semantic Cloud of Tags*) est un exemple du premier type d'approches. Ce programme a pour but la description de la structure et des relations, dites sémantiques, d'un ensemble de « *tags* » afin d'en favoriser les usages socio-cognitifs, et ce en s'appuyant sur le fait que les processus de « *tagging* » mettent en jeu des personnes (acteurs), des ressources et des « *tags* ». D'un point de vue général, ainsi que nous l'avons déjà indiqué et en appui sur une partie des analyses concernant les réseaux sociaux, le cadre conceptuel et les outils du Web Sémantique prennent en charge la représentation et l'échange des savoirs au sein des réseaux sociaux numériques, grâce à RDF, au langage de requête SPARQL[33] à RDFS et OWL. Les folksonomies peuvent en effet être exploitées en utilisant la théorie des graphes, et ce en vue d'identifier les groupes d'utilisateurs et les thématiques émergentes majeures. D'autres recherches ont étendu le langage SPARQL dans le but de mettre à jour les connexions et les chemins de type sémantique entre des ressources RDF.

Il convient toutefois de noter l'exploration de nouvelles pistes concernant ce que l'on peut appeler des « ontologies du tag ». L'intérêt de ce type d'ontologie est, parmi d'autres caractéristiques, d'« intégrer l'aspect "actes de langages" » associé aux tags (ce que l'on fait en taguant), de permettre à

n'importe quelle ressource de faire office de tag, de réutiliser les travaux menés pour résoudre l'"identity crisis" du Web sémantique, enfin de souligner la dimension sémiotique de la relation entre le tag et la ressource (non plus ce que l'on fait du signe, mais ce que le signe fait en tant que signe – les deux dimensions s'ajoutant l'une à l'autre) ». [34]

### G. Alternatives

D'autres conceptions essaient de prendre en compte les pratiques communicationnelles « associés à la conduite d'interactions éphémères entre utilisateurs distants tout en offrant des représentations, souvent de nature graphique, des réseaux sociaux ainsi constitués » Elles défendent une conception pragmatique des processus informationnels et communicationnels, et envisage la linguistique et la sémiotique de manière plus ouverte. Il y a, selon nous, un grand intérêt à ne pas laisser le champ libre au seul formalisme évoqué par le « cake » [35] de Tim Berners-Lee, dont l'efficacité est subordonnée à une « fermeture sémiotique », c'est-à-dire à une réduction et à une standardisation des comportements et des pratiques.

Quelle est, là encore la position de ceux qui parlent et agissent au nom de l'Open Data et souhaitent ouvrir la dissémination de nouvelles pratiques sociales, intellectuelles adaptées ?

Il est possible en effet de proposer d'autres postures prenant en compte la sociologie en acte des pratiques et des usages, ainsi que les phénomènes de co-construction des connaissances, afin que « les langages syntaxiquement formels » puissent précisément être efficaces. Le problème est de concevoir des méthodes qui puissent représenter de telles structures sémiotiques et socio-cognitives, de manière à ce qu'un formalisme faible rende possible, à travers de nouvelles écritures, une approche pragmatique forte. C'est la raison pour laquelle il nous semble important de discuter, de manière critique, l'élaboration de ces nouveaux alphabets, de leurs contraintes combinatoires et de leurs grammaires. Il convient également de réfléchir à des nouvelles manières « non-documentaires » de produire des onto-éthologies [36] ouvertes et dynamiques, et ce afin de se rappeler que les écritures s'évaluent et s'imposent à partir de ce qu'elles ouvrent de créativité et d'inventions, de ce qu'elles portent de nouveaux modes combinatoires comme autant d'herméneutiques possibles.

Ces modes combinatoires sont particulièrement visibles dans le cas de traitements des folksonomies qui utilisent des approches statistiques et des méthodes issues par exemple de l'analyse des mots associés.[37] Ces méthodes, qui sont fondées sur la similarité et le principe de co-occurrence, montrent que les pratiques de « taggage » présentent des caractères proches de ceux de l'indexation traditionnelle et

que ces caractères sont relativement prédictibles. [38] De plus, la multiplication des folksonomies, leur hétérogénéité et leur caractère non-interopérable poussent à analyser, selon la même philosophie, les phénomènes de chevauchement entre « espaces de tags », en particulier dans les espaces collaboratifs. [39]

Plutôt que des ontologies, il faut donc pouvoir aussi accéder à la définition des « onto-éthologies ». [40] Elles expriment les structures socio-cognitives portées par les corpus, les traductions et les processualités à l'œuvre au cœur même des communautés. La « structuration » (formalisation) des textes et des documents, de même que leur filtrage, doivent être envisagés, dans leurs aspects techniques, sous une double contrainte. Il faut pouvoir traiter des populations de textes numériques susceptibles d'être en permanence recomposées et transformées, d'une part ; il faut fabriquer des outils d'exploration et d'exploitation socio-cognitives, intellectuelles de ces populations, des outils de représentation de leurs processualités constitutives qui favorisent des capacités analogiques, associationnistes et combinatoires, selon des niveaux d'organisation multiples, d'autre part. L'Open Data et son insertion créative est au prix de cette accroissement de réflexivité qui découle selon nous de telles possibilités et des nouvelles pratiques cartographiques émergentes associées, qu'elles soient doxiques ou savantes.

#### *H. Que signifie l'expression ouvrir les données ? L'exemple de l'Open Archive*

A partir des corpus numériques d'une communauté de recherche, les logiciels de traitements automatiques et semi automatiques des données (dans le cas présent, des collections d'articles ou de documents produits par les acteurs de la recherche) évoqués ci dessus, peuvent représenter les associations, les réseaux d'association, les modes d'agrégation et de sélection, les contraintes et les modes combinatoires, les modes sociaux de transmission-sélection de ces contraintes. Ils sont à l'œuvre dans les agencements hétérogènes des chercheurs, des laboratoires, des textes, des revues, des thématiques ou des concepts. Ouvrir les données pour de « nouvelles visibilités » implique également d'être capable de mettre à jour les fronts de recherche, les réseaux d'influence, les systèmes de traduction, de chevauchement ou de percolation des notions, des concepts ou des thèmes. Les représentations doivent inclure les réseaux d'acteurs, les réseaux de citations, les co-citations[41] et les modes de répétition ou d'altération des textes et des contextes associés, ou encore les graphes conceptuels. La finalité est d'offrir aux chercheurs de nouvelles façons de s'orienter, et donc d'amener à une meilleure gestion-navigation des points de vue, à augmenter les capacités associationnistes de ce qui constitue, pour partie, nos conditions structurales de visibilité, lesquelles seront toujours singulières et bornées. C'est là le sens profond de ce que l'on appelle les nouvelles pratiques cartographiques,

qui devraient alors susciter de nouveaux types d'interfaces. Dans une perspective proche, les recherches portées par le courant « memétique » [42] proposent une conception des modes de répliques et de vie des « memes », conçus comme des entités sémiotiques, des formes et des structures variables. Que signifie donc pour les collectivités territoriales et ceux qui accéderont aux données, « ouvrir les données pour de nouvelles visibilités » ? Quels types de réflexivité seront alors introduites qui affecteront les usages et les pratiques socio-cognitives ?

Ces approches soulèvent bien des questions, dès lors qu'elles semblent parfois s'appuyer sur une conception parfois par trop essentialiste des interactions et des usages, alors qu'elles visent la vie de populations, de multiplicités d'agencements comme... multiplicités métastables.

Cette réflexion vise à attirer l'attention sur les diverses manières d'envisager les problèmes de description et de représentation des mémoires numériques hypertextuelles, de créer du sens à partir d'elles et de montrer la nécessaire pluralité quant à la manière de concevoir – suivant les contextes et les corpus, les pratiques et les usages –, les types d'écritures dont nous avons besoin pour habiter les multiples espace-temps socio-cognitifs. Les différentes approches de structuration du Web ne sont pas fondamentalement incompatibles et les manières dont nous pourrions développer les usages des Open Data et leur Désir est à ce prix.

Nous sommes donc davantage favorables à une approche ouverte qui consisterait à penser les rapports différentiels entre des conceptions différentes du Web sémantique à mieux en définir les articulations et les niveaux d'articulation.

Une telle position laisse la vie des dissensus au centre de l'habitat démocratique comme condition essentielle de sa perpétuation et de son approfondissement.

#### REFERENCES

- [1] Yann Moulier-Boutang « Droits de propriété intellectuelle, terra nullius et capitalisme cognitif », *Multitudes* 2/2010 (n° 41), p. 66-72.
- [2] Voir les Plateaux 12 et 23 in G. Deleuze et F. Guattari, *Capitalisme et schizophrénie 2*, Edition de minuit, 1980 - Voir aussi Antonio Negri, Michael Hardt, *Empire*, Harvard University Press, New Ed, 2001
- [3] Yann Moulier-Boutang « Droits de propriété intellectuelle, terra nullius et capitalisme cognitif », *Multitudes* 2/2010 (n° 41), p. 66-72.
- [4] [http://fr.wikipedia.org/wiki/Lawrence\\_Lessig](http://fr.wikipedia.org/wiki/Lawrence_Lessig); <http://www.stallman.org/>; <http://www.openarchives.org/>; <http://archivistic.ccsd.cnrs.fr/>; <http://arxiv.org/>
- [5] <http://opendata.paris.fr/opendata/jsp/site/Portal.jsp>; <http://www.datalocale.fr/thematiques> – <http://www.data.rennes-metropole.fr/...>
- [6] [http://fr.wikipedia.org/wiki/Open\\_Government\\_Initiative](http://fr.wikipedia.org/wiki/Open_Government_Initiative) [http://en.wikipedia.org/wiki/Open\\_source\\_governance](http://en.wikipedia.org/wiki/Open_source_governance) <http://www.whitehouse.gov/open>
- [7] JM Noyer, B Juanals, La stratégie américaine du contrôle continu De la « Noopolitik » (1999) à « Byting Back » (2007) : une création de concepts et de dispositifs de contrôle des populations ([http://archivistic.ccsd.cnrs.fr/sic\\_00292207/fr](http://archivistic.ccsd.cnrs.fr/sic_00292207/fr), voir aussi Counterinsurgency FM 3-24, David H. Petraus, James F. Amos, respectivement Lieutenant General, U.S. Army Lieutenant General,

- U.S. Marine Corps Commander Deputy Commandant U.S. Army Combined Arms Center Combat Development and Integration).
- [8] Sur ce point voir le travail de Stephen Graham « Cities under Siege. The New Military Urbanism », 2010
- [9] , JM Noyer, La transformation numérique: quelques procès en cours : Narratique, Empire, Entreprise, Data mining, Formes courtes, A paraître in Revue Française des Sciences de l'Information et de la Communication, Publisher, Sfsic. <http://www.openedition.org/10877>, [http://hal.archives-ouvertes.fr/index.php?halsid=r0to0cn44trk8db4lp37v1b5&view\\_this\\_doc=sic\\_00725836&version=1](http://hal.archives-ouvertes.fr/index.php?halsid=r0to0cn44trk8db4lp37v1b5&view_this_doc=sic_00725836&version=1)
- [10] P. Mengue, Utopies et devenirs deleuziens, Ed. L'Harmattan, 2009
- [11] JM Noyer et M Carmes, Les interfaces machiniques comme problème sémio-politique, in « Enjeux politiques du document numérique », « Document Numérique et Société », ADBS éditions, 2011, chap.11, pp193-216 et Le « pluralisme sémiotique » et le nouveau gouvernement des signes, Hommage à Félix Guattari, Maurizio Lazzarato, 2006
- [12] Gilles Deleuze, Felix Guattari, Mille plateaux, Editions de minuit, 1981
- [13] Louise Merzeau, <http://archivesic.ccsd.cnrs.fr/docs/00/48/32/94/PDF/CIT-ES-MERZEAU.pdf>
- [14] Antonio Negri, Michael Hardt, Empire, Harvard University Press, New Ed, 2001
- [15] Le paradigme de Jim Gray: ses paramètres, ses conséquences. The Fourth Paradigm, Data-Intensive Scientific Discovery. GORDON BELL (Fourth's Paradigm). Tony Hey, Stewart Tansley, Kristin Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery, 2009. Voir aussi la discussion des thèses de Chris Anderson (Editor-in-Chief of *Wired* magazine) The end of theory: Will the Data Deluge Makes the Scientific Method Obsolete? And On Chris Anderson's The end of Theory, by George Dyson, Kevin Kelly, Stewart Brand, W. Daniel Hillis, Sean Carroll, Jaron Lanier, Joseph Traub, John Horgan, Bruce Sterling, Douglas Rushkoff, Oliver Morton, Daniel Everett, Gloria Origgi, Lee Smolin, Joel Garreau. Voir aussi, Les Dossiers de la Recherche: La révolution des mathématiques, en particulier "L'informatique renouvelle les mathématiques" par Gerard Berry, Directeur de recherche à l'Inria, et membre de l'académie des sciences.
- [16] Chris Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, in *Wired* magazine, 16.07, 2008
- [17] JM Noyer, B Juanals, La question pragmatique dans le contexte des mémoires numériques, H2PTM'07, Edition Hermès Lavoisier, 2007
- [18] Jean-François Lyotard, La condition Postmoderne, Editions de Minuit, 1979
- [19] Yannick Maignien, Les nouvelles frontières numériques des sciences (Editorial, Adonis 2009)
- [20] <http://poynder.blogspot.fr/2012/06/new-declaration-of-rights-open-content.html> et en particulier l'interview de Peter Murray-Rust par Richard Poynder (2012)
- [21] In, La transformation numérique: quelques procès en cours Narratique, Empire, Entreprise, Data mining, Formes courtes... Jean-Max Noyer, [http://archivesic.ccsd.cnrs.fr/sic\\_00725836/fr/](http://archivesic.ccsd.cnrs.fr/sic_00725836/fr/), à paraître dans la Revue Française des Sciences de l'Information et de la communication (2013)
- [22] [http://fr.wikipedia.org/wiki/Données\\_ouvertes\\_en\\_France](http://fr.wikipedia.org/wiki/Données_ouvertes_en_France)
- [23] : Il y a là un champ d'investigation très intéressant. Les diverses pratiques d'insertion et création de métadonnées dans les divers documents et corpus... selon les secteurs d'activité et les types de documents, ne sont pas suffisamment étudiées
- [24] E.Ostrom, Beyond markets and states: polycentric governance of complex economic systems, Nobel lecture, December 8, [http://www.w3.org/2011/gld/wiki/Main\\_Page](http://www.w3.org/2011/gld/wiki/Main_Page)
- [25] [http://www.w3.org/2011/gld/wiki/Main\\_Page](http://www.w3.org/2011/gld/wiki/Main_Page)
- [26] Pierre Lévy « Au-delà de Google », *Multitudes* 1/2009 (n° 36), p. 45-52. URL : [www.cairn.info/revue-multitudes-2009-1-page-45.htm](http://www.cairn.info/revue-multitudes-2009-1-page-45.htm).
- [27] Philip K. Dick, "The Minority Report", 1956. Thomas Berns : « L'efficacité comme norme » in Dissensus, Revue de philosophie politique de l'ULG – N°4 – Avril 2011. Voir aussi le concept de préemption dans le domaine stratégique. Bernard E. Harcourt, Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age. Chicago, The University of Chicago Press, 2007.
- [28] <http://www.ieml.org/french/elements.html> <http://archivesic.ccsd.cnrs.fr/docs/00/06/87/32/PDF/ieml-v2.pdf>
- [29] Douglas Hofstadter, Godel, Escher, Bach, Les brins d'une guirlande éternelle, Editions Interéditions, 1985 et Voir, aussi, Ontologies pour le Web sémantique, Jean Charlet, Bruno Bachimont, Rapahél Troncy spécifique 32 CNRS / STIC, Web sémantique, Rapport final, Éditeurs Jean Charlet, Philippe Laublet & Chantal Reynaud V3 – décembre 2003, Charlet .
- [30] OAI-ORE est conçu pour surmonter les barrières entre entrepôts et pour développer une nouvelle génération de fonctionnalités et d'outils transversaux. À partir de ce point de vue, OAI-ORE propose de standardiser la description des relations entre les documents numériques. Par relations entre documents, il est entendu, par exemple, les relations entre différentes versions d'un document, ou bien, l'aggrégation de documents, telle qu'une page web avec ses images ou un ensemble de chapitres dans un ouvrage. En ligne : <http://www.openarchives.org/ore/1.0/datamodel>
- [31] ZACKLAD M., « Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information », in actes de la Conférence CAIS/ACSI 2007, Montréal, [http://www.cais-acsi.ca/search\\_fr.asp?year=2007](http://www.cais-acsi.ca/search_fr.asp?year=2007)
- [32] Méthodes des mots associés et calcul de fréquences de co-occurrences des termes, par exemple. Voir sur ces questions : Courtial J.-P., H. Penan, M. Callon (dir.), *Introduction à la Scientométrie*, PUF, 1990. Revue Solaris n° 2, dossier « Bibliométrie, Scientométrie », 1995. <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/index.html>
- [33] Le langage SPARQL définit la syntaxe et la sémantique nécessaire à l'expression de requêtes sur une [base de données](#) de type [RDF](#) et la forme possible des résultats. SPARQL est adapté à la structure spécifique des graphes [RDF](#) et s'appuie sur les triplets qui les constituent. Voir en ligne : <http://www.w3.org/2001/sw/>. RDF Schema ou RDFS est un langage extensible de [représentation des connaissances](#). RDFS fournit des éléments de base pour la définition d'[ontologies](#) ou vocabulaires destinés à structurer des ressources [RDF](#). Web Ontology Language (OWL) est un [dialecte XML](#) basé sur une syntaxe [RDF](#). Il fournit les moyens pour définir des [ontologies web](#) structurées. Le langage OWL est basé sur la recherche effectuée dans le domaine de la [logique de description](#).
- [34] Communication d'Alexandre Monin : ontologie proposée par David Laniado de l'École Polytechnique de Milan, Freddy Limpens de l'INRIA, Alexandre Monin, assistés de Thomas Lörtsch et Ricardo Tasso. Voir aussi le compte rendu du VoCamp organisé par l'Inria, Septembre 2009 <http://vocamp.org/wiki/VoCampNiceSeptembre2009>.
- [35] L'infrastructure dédiée au Web Sémantique est souvent présentée sous la forme d'un « Cake », dit de Tim Berners-lee.
- [36] Par éthologie conceptuelle, nous entendons le champ, le diagramme à partir desquels émergent les concepts comme singularités étant tissées d'autres concepts, de percepts, d'affects, d'écritures plus ou moins hétérogènes, de contraintes combinatoires diverses. Voir Alliez É, *La signature du monde ou qu'est-ce que la philosophie de Deleuze-Guattari*, Paris, Éd. du Cerf, 1993. Plus précisément le chapitre III, Onto-éthologies.
- [37] Ciro Cattuto, Dominik Benz, Andreas Hotho, Gerd Stumme, « Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems », arXiv :0805.2045v1, 2008. Research on the semantic-based co-word analysis, Zhong-Yi Wang • Gang Li • Chun-Ya Li • Ang Li, Akadémiai Kiadó, Budapest, Hungary 2011. D'un point de vue plus général voir l'ensemble des travaux issus de l'analyse des mots associés et des approches fractales de la co-construction des connaissances. Et Collaboration and cognitive structures in social science research fields. Towards socio-cognitive analysis in information systems, Peter Mutschke, Anabel Quan Haase, Budapest Scientometrics, Vol. 52, No. 3 (2001) 487–502
- [38] Margaret E.I. Kipp and D. Grant Campbell, « Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices », 2008, ACM.
- [39] *Analysis of Tag / Co-Tag Spaces in Social Classification Systems*, Ed. Springer-Verlag, 2009.
- [40] Cf. note 37
- [41] Les questions portées par la citation sont complexes. Nous renvoyons à Jacques Derrida, en particulier à, *Limited Inc.*, Éd. Galilée, 1990, et à Antoine Compagnon, *La Seconde Main ou le travail de la citation*, Paris, Éd. Seuil, 1979. Voir aussi : Betsy Van der Veer Martens, "Do Citation Systems Represent Theories of Truth?", School of Information

- Studies, Syracuse University. Case, D.O and G.M. Higgins : “How Can We Investigate Citation Behavior? A Study of Reasons for Citing Literature in Communication.” *Journal of the American Society for Information Science*, 2000, 51. Wouters, P. : “Beyond the Holy Grail: From Citation Theory to Indicator Theories.” *Scientometrics*, 44, 1999. Cronin, B. “Metatheorizing Citation”, *Scientometrics*, 43,1998. Leydesdorff, Loet, “Theories of Citation? », *Scientometrics*, 43, 1998.
- [42] Journal of Memetics – Evolutionary Models of Information Transmission, <http://cfpm.org/jom-emit/>  
<http://pespmc1.vub.ac.be/MEMES.html>  
<http://en.wikipedia.org/wiki/Memetics>